

Customer Lifetime Value (CLV):

Customer Lifetime Value (CLV) is the total revenue a business can expect from a customer over the entire duration of their relationship. It helps businesses understand how valuable a customer is beyond just a single purchase.

Importance:

- **Optimize Marketing Strategies** – Identify high-value customers and invest in retaining them.
- **Improve Customer Retention** – Focus on customers who will generate the most revenue long-term.
- **Personalize Offers** – Provide tailored discounts or loyalty programs based on CLV.
- **Budget Allocation** – Determine how much to spend on acquiring new customers vs. retaining existing ones.
- **Business Growth & Forecasting** – Estimate future revenues and make data-driven decisions.

Dataset preparation:

- Average Order Value (AOV) = Total Revenue / Total Orders
- Purchase Frequency (PF) = Total Orders / Total Customers
- Recency (R) = Days since last purchase

These are calculated fields to build ML model.

Required structure of dataframe

	customer_id	total_amount	order_id	AOV	order_date	recency
0	101	1400	4	350.000000	2023-10-05	76
1	102	1050	3	350.000000	2023-11-05	45
2	103	1350	3	450.000000	2023-11-12	38
3	104	1650	3	550.000000	2023-12-01	19
4	105	1070	3	356.666667	2023-12-10	10
5	106	870	3	290.000000	2023-12-15	5
6	107	1350	3	450.000000	2023-12-20	0
7	108	650	2	325.000000	2023-09-05	106

CLV = AOV * frequency

	customer id	AOV	order freq	recency	clv
0	101	350.000000	4	76	1400.0
1	102	350.000000	3	45	1050.0
2	103	450.000000	3	38	1350.0
3	104	550.000000	3	19	1650.0
4	105	356.666667	3	10	1070.0
5	106	290.000000	3	5	870.0
6	107	450.000000	3	0	1350.0

Building ML model:

Preprocessing:

1. Load data
2. Check datatypes
3. Converting datatype to required datatype if needed
4. Checking nulls, infinite values, duplicates

Building model:

5. Devide and assign features and targets
feature = x = 'AOV', 'order_freq', 'recency'
target = y = clv
6. Dividing testing and training part, testing is 30% and traing is 70% of total data
7. Make predictions by running model
8. Check the accuracy of model:

RMSE: 54.49 – this depends upon range of CLV.

R² Score: 0.97 - 97% of the variability in CLV is explained by your features (aov, or der_freq, recency). The model fits the data well